

IN THE CLAIMS:

The text of all pending claims, (including withdrawn claims) is set forth below. Cancelled and not entered claims are indicated with claim number and status only. The claims as listed below show added text with underlining and deleted text with ~~strikethrough~~. The status of each claim is indicated with one of (original), (currently amended), (cancelled), (withdrawn), (new), (previously presented), or (not entered).

Please AMEND claims 1-4, 16, 19 and 21, and ADD claim 22 in accordance with the following:

1. (currently amended) A method for collecting documents linked to each other from a network by crawling the network, comprising:

collecting first a number of documents into a collected document group until the number of documents is equal to or larger, in number, than a predetermined value, the documents being collected from inside a community through the network based on a reference relation of each document in the first documents collected document group, the reference relation defining a relationship between the first-referring documents and second-referenced documents inside or outside the community which are linked to the first-referring documents; and

collecting third documents a newly collected document group from inside and outside the community based on the reference relation of the first documents after collecting the first documents collected document group from inside the community.

2. (currently amended) The method according to claim 1, further comprising:

computing a significance level indicating a level of significance of a collected document belonging to the newly collected document group according to the reference relation of references to the collected document, and information about a position of the collected document in the network; and

determining a new documents to be collected based on the reference relation the references from and the significance level of the collected document.

3. (currently amended) The method according to claim 2, wherein ~~said collecting of the third-new documents is~~ are determined separately for inside the community and for outside the community.

4. (currently amended) The method according to claim 3, further comprising:
presenting a result of retrieving the ~~third~~new documents separately for inside the community and outside the community.
5. (previously presented) The method according to claim 2, further comprising:
determining whether any of the documents is in the community according to information indicating the position of the document in the network.
6. (cancelled)
7. (previously presented) A method for collecting documents linked to each other from a network by crawling the network, comprising:
providing a positive sample document group relating to a field, and a negative sample document group related to the field;
determining a prospective document for collection that is related to the field based on a reference relation to the positive sample document group and the negative sample document group by computing a reference score indicating a level at which the prospective document is referenced only by at least one document in the positive sample document group based on the reference relation which defines a relationship between original documents and linked documents belonging to the positive sample document group or to the negative sample document group which are linked to the original documents; and
collecting the prospective document having a high reference score.
8. (previously presented) A method for collecting documents linked to each other from a network by crawling the network, comprising:
providing a positive sample document group relating to a field, and a negative sample document group less related to the field;
determining a prospective document for collection that is related to the field based on a reference relation to the positive sample document group and the negative sample document group by computing a co-reference score indicating a level at which the prospective document is referenced together with a first collected document in the positive sample document group for a second collected document referenced by a third collected document referring to a fourth collected document in the positive sample document group based on the reference relation which defines a relationship between original documents and linked documents belonging to the

positive sample document group or to the negative sample document group which are linked to the original documents; and

collecting the prospective document having a high co-reference score.

9. (cancelled)

10. (previously presented) The method according to claim 1, further comprising:
summarizing the collected documents based on a referencing expression used in the collected documents.

11. (previously presented) The method according to claim 1, further comprising:
assigning a keyword to the collected documents based on a referencing expression used in the collected documents.

12. (original) The method according to claim 1, further comprising:
not assigning a keyword based on the referring expression when the referencing expression is used regardless of a content of a referenced document.

13. (original) The method according to claim 11, further comprising:
counting a number of different documents referenced using the referencing expression;
and
not assigning the keyword based on the referencing expression when the number of different documents is equal to or larger than a predetermined value.

14. (original) The method according to claim 11, further comprising:
counting a reference frequency at which each collected document is referenced by the referencing expression when the number of different documents is smaller than a predetermined value; and
determining whether or not the referencing expression is assigned as the keyword based on the number of different documents and the reference frequency.

15. (original) The method according to claim 11, further comprising:

combining the keyword based on the referencing expression with a keyword extracted from text of the collected document, and a keyword extracted from information indicating a position in the network about the collected document.

16. (currently amended) A method for retrieving documents ~~linked to each other~~ from a terminal belonging to a community in a network ~~by crawling the network~~, comprising:

transmitting information for retrieval of the documents to a server; and

receiving the documents retrieved separately from inside and outside the community according to the information for retrieval together with information indicating a significance level for the community, the significance level ~~indicating a relationship between original documents and other documents inside or outside the community which are linked to the original documents~~ an importance of a document based on information about references in a collected document group, and information indicating a position of the document.

17. (previously presented) A document collection apparatus collecting documents linked to each other from a network by crawling the network, comprising:

a next prospect determination unit determining a prospect to be collected next based on a reference relation of a collected document, the reference relation defining a relationship between original documents and other documents inside or outside the community which are linked to the original documents;

a community determination unit determining whether or not the prospect is in a community in the network according to information indicating a position in the network of the prospect; and

a document collection unit collecting the prospect from the network, wherein said document collection unit collects the prospect from inside and outside the community after collecting documents larger in number than a predetermined value from inside the community.

18. (previously presented) A document collection apparatus collecting documents linked to each other from a network by crawling the network, comprising:

a next prospect determination unit determining a prospect to be collected next based on a reference relation between a positive sample document group which is a document group related to a field and a negative sample document group which is a document group less related to the field; and

a document collection unit collecting the prospect from the network; wherein the reference relation is a relationship between original documents and other documents belonging to the positive sample document group or to the negative sample document group which are linked to the original documents.

19. (currently amended) A computer-readable recording medium recording a program used to direct a computer to control collection of documents linked to each other from a network by crawling the network, comprising:

collecting ~~first~~ a number of documents into a collected document group until the number of documents is equal to or larger, in number, than a predetermined value, the documents being collected from a community through the network based on a reference ~~relation of each document in the first documents collected document group~~, the reference ~~relation~~ defining a relationship between ~~the first-referring documents and second-referenced documents~~ inside or outside the community which are linked to the first-referring documents; and

collecting ~~third documents~~ a newly collected document group from inside and outside the community based on the reference ~~relation of the first documents after said collecting of the first documents equal to or larger, in number, than the predetermined value~~ collected document group from inside the community.

20. (previously presented) A computer-readable recording medium recording a program used to direct a computer to control collection of documents linked to each other from a network by crawling the network, comprising:

providing a positive sample document group which is a document group relating to a field, and a negative sample document group which is a document group less related to the field;

determining a document to be collected relating to the field based on a reference relation to the positive sample document group and the negative sample document group, the reference relation defining a relationship between original documents and other documents inside or outside the community which are linked to the original documents; and

collecting the document to be collected from the network.

21. (previously presented) A computer data signal embodied on a carrier expressing a program used to direct a computer to control collection of documents linked to each other from a network by crawling the network, said program instructing the computer to perform the process comprising:

collecting ~~first-a number of documents into a collected document group until the number of documents is equal to or larger, in number, than a predetermined value, the documents being collected~~ from a community through the network based on a reference ~~relation-of each document in the first documents collected document group~~, the reference ~~relation~~-defining a relationship between the ~~first-referring~~ documents and ~~second-referenced~~ documents inside or outside the community which are linked to the first-referring documents; and

collecting ~~first-a number of documents into a collected document group until the number of documents is equal to or larger than, in number, a predetermined value, the documents being collected~~ from inside a community in the network based on a reference ~~relation-of each document in the first documents collected document group~~, the reference ~~relation~~-defining a relationship between the ~~first-referring~~ documents and ~~second-referenced~~ documents inside or outside the community which are linked to the ~~first-referring~~ documents; and

collecting ~~third-documents-a newly collected document group~~ from inside and outside the community based on the reference ~~relation-of the first documents after said collecting-of the first documents equal to or larger, in number, than the predetermined value-collected document group~~ from inside the community.

22. (new) A method for collecting documents linked to each other from a network by crawling the network, comprising:

collecting a number of documents into a collected document group until the number of documents is equal to or larger than a predetermined value, the documents being collected from inside a community through the network based on references in the documents in the collected document group, each reference defining a relationship between a referring document and a referenced document, inside or outside the community, which is linked to the referring document; and

collecting a newly collected document group from inside and outside the community based on the reference of the collected document group from inside the community.